Published in final edited form as:

Trends Cell Biol. 2018 October ; 28(10): 764–775. doi:10.1016/j.tcb.2018.05.004.

Analyzing circulating tumor cells one at a time

Veronica Ortiz^{a,b} and Min Yu^{a,b,*}

^aDepartment of Stem Cell Biology and Regenerative Medicine, University of Southern California, Los Angeles, CA 90033, USA

^bUSC Norris Comprehensive Cancer Center, Keck School of Medicine, University of Southern California, Los Angeles, CA 90033, USA

Abstract

Whole-genome sequencing has made a significant impact on cancer research, but traditional bulk methods fail to detect information from rare cells. Recently developed single-cell sequencing methods have provided new insights and unprecedented details about cancer progression and diversity. These advancements also enable the investigation of rare cells, such as circulating tumor cells (CTCs) derived from cancer patients. In this review, we outline various single-cell sequencing techniques that can elucidate the molecular properties of CTCs. In addition, we explain the drawbacks that need to be overcome for each method.

Keywords

Circulating Tumor Cells; Single-Cell Sequencing; Tumor Heterogeneity

Single-Cell Sequencing of Circulating Tumor Cells

Over the past few years, whole-genome sequencing has significantly advanced our understanding of cancer. However, many of these whole-genome sequencing methods require a high input of DNA or RNA from numerous cells, and ultimately depict an average of a complex population. Consequently, these methods have limitations in investigating both rare individual cells and precious samples containing only a few cells. To address these problems, many recently developed single-cell sequencing methods have facilitated our understanding of the progression and diversity of cancer at the single-cell level. These advancements make it possible to study precious cells derived from cancer patients, such as circulating tumor cells (CTCs). CTCs are tumor cells that originate from either primary or metastatic tumors and travel through systemic circulation to distant organs, where they can initiate metastatic lesions [1]. Additionally, it has been shown that CTCs can re-seed and colonize their tumors of origin, thus adding to the heterogeneous nature of these tumors [2]. CTCs can be isolated from patient blood samples, potentially enabling an early detection of

*Correspondence: minyu@med.usc.edu (M. Yu).

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

metastatic progression, therapeutic response and tumor relapse [3]. Due to intra-tumor heterogeneity, traditional surgical biopsies that are taken from one part of a tumor can miss information contained in other active regions [4]. In contrast, CTCs can consist of a mixture of cells shed from multiple active tumor regions, potentially providing a better representation of the invasive clones. In addition, because of the minimally invasive procedure of isolating CTCs, it is a more practical approach for repeatedly monitoring disease progression [5].

Over the past decade, techniques of CTC isolation have improved [3, 6], especially due to our growing understanding of CTCs and technological advancements. Many of these techniques focus on the positive selection of CTCs, via surface markers present on tumor cells or through negative selection that aims at eliminating white blood cells (WBCs) and red blood cells (RBCs). However, the rarity and delicacy of CTCs still makes downstream analysis a challenge. It is estimated that in most metastatic patients there are fewer than 10 cells per mL of blood, with a mixture of 1 million WBCs and 1 billion RBCs [7]—and only about half of these CTCs are viable in circulation [8, 9]. Thus, single-cell sequencing methods have become essential for investigating these rare cancer cells. Single-cell methods enable us to investigate how the heterogeneity of a CTC population contributes to therapeutic resistance [10]. In addition, it is possible that the comparison of single CTCs to the primary and metastatic tumors can aid in the discovery of metastasis-initiating clones and can reveal molecular biomarkers for metastasis. However, such an analysis requires the consideration of the different molecular characteristics of CTCs, such as genetics, epigenetics, and the transcriptome (Figure 1, Key Figure). In this review, we outline various single-cell sequencing techniques (Box 1) and explain how these techniques can expand our knowledge of the diverse characteristics of CTCs. In addition, we explain the drawbacks that need to be overcome for each method. However, many analyses have not been applied to CTCs, thus we will describe from the most established to the most novel techniques.

Box 1

Methods for analyzing the CTC genome

RNA Sequencing (RNA-seq) [74]: Reveals the presence and quantity of RNA transcripts by sequencing the whole transcriptome from a particular cell or cells. Through this method, total RNA is converted to cDNA, which is later used as an input for next-generation sequencing (NGS) after library preparation.

DNA Sequencing (DNA-seq): Determines the precise order of nucleotides within the genome. DNA-seq can be used to determine mutations or copy number variations within the genome. DNA is used as the input for NGS sequencing after library preparation.

Whole Genome Bisulfite Sequencing (WGBS) [48, 49]: Detects methylated cytosines within the genome and is currently considered the gold-standard for detecting DNA methylation. DNA is treated with sodium bisulfite, which chemically changes unmethylated cytosines to uracils. Through amplification and NGS, these uracils are changed to thymines which enables the detection of methylation up to the single nucleotide level. Sodium bisulfite treated DNA is used as an input for NGS after library preparation.

DNase I hypersensitive Sequencing (DNase-seq) [75]: Detects accessible regions in the genome via sequencing after DNase I treatment. DNA is treated with DNase I, which cuts at sites that are sensitive to cleavage. Regions of DNA that are affected by DNase I manifest regions that are exposed and assessible, thus pertaining to potential active promoters or other regulatory regions within the genome. Treated DNA is then used as an input for NGS after library preparation.

Assay for Transposase-Accessible Chromatin Sequencing (ATAC-seq) [76]: Determines chromatic accessibility via sequencing after adapter addition by transposase Tn5 on exposed DNA. The regions affected by Tn5 are exposed and accessible, thus identifying potential active promoters or other regulatory regions within the genome. Tn5 treated DNA is then used as an input for NGS after library preparation.

Hi-C Sequencing [77]: Examines three-dimensional (3D) architecture of whole genomes. It is mostly used to determine the interactions between genomic loci in 3D space, such as promoter-enhancer interactions. Chromatin is crosslinked with formaldehyde, then digested, and later ligated to determine the 3D organization of DNA. Digested and ligated DNA is used as an input for NGS after library preparation.

Chromatin Immunoprecipitation Sequencing (ChIP-seq) [78]: Examines protein interactions with DNA. DNA and associated proteins in cells are crosslinked, sheared, and immunoprecipitated with antibodies against specific proteins. De-crosslinked DNA fragments are then used as an input for NGS after library preparation.

In situ hybridization (RNA-ISH) [16]: Profiles the transcriptome. Cells are fixed and mRNA is hybridized by probes within intact cells. The intensities of labeled probes are amplified by other rounds of hybridization, are localized and quantified within intact cells by using fluorecence microscopy.

RNA sequential probing of targets (RNA SPOTs) [23]: Profiles mRNAs with single-molecule sensitivity based on sequential fluorescent in situ hybridization (FISH). mRNAs are hybridized by probes that contain combinations of unique pseudocolors, which are detected by rounds of barcoding and imaging. Its multiplexing capacity can detect over 10,000 mammalian genes, with potential advantages of additional spatial information and is applicable for lowly transcribed genes.

Transcriptome

Analyzing the changes of total transcribed genes in a cell, termed transcriptome, is probably one of the most commonly used applications of next generation sequencing. By studying the transcriptome, we can reveal different variants and expressional levels of genes. Genome-wide transcriptional studies have been beneficial in understanding the molecular mechanisms involved in prognosis and drug resistance within cancer. Data from The Cancer Genome Atlas (TCGA) has provided extensive evidence of inter-patient heterogeneity in gene expression profiles from almost all cancer types, thus providing a rationale for patient stratification for clinical management (<https://cancergenome.nih.gov>). However, most of these TCGA studies represent averages of the gene expression profiles from populations of

cells, and thus lack information about intra-tumor heterogeneity. Intra-tumor heterogeneity plays a crucial role in disease progression under selective pressures, such as the multiple steps of the metastatic cascade and various therapies [11]. Due to the difficulties of retrieving repeated surgical biopsies from multiple sites, analyzing expression patterns of CTCs at the single-cell level can provide insights into molecular pathways altered during metastasis and the evolution of a heterogeneous population during therapy resistance.

Full length single-cell RNA-sequencing (scRNA-seq) was first applied to six CTCs from a melanoma patient and the data was compared with various cell types, such as primary melanocytes, melanoma cancer cell lines, human embryonic stem cells (ESCs), and immune cells [12]. The scRNA-seq analysis differentiated CTCs from other cell types and uncovered signaling pathways altered in CTCs relative to melanocytes, such as an upregulation of melanoma associated antigens and a downregulation of cell death and MHC class I genes. In addition, this study also identified an upregulation of several genes encoding plasma membrane-associated proteins and a downregulation of genes associated with escaping immune surveillance. These upregulated surface markers can potentially differentiate CTCs from primary melanocytes and blood cells, which can facilitate future CTC isolation in melanoma patients [8]. In order to identify metastasis-promoting signals, we have demonstrated that the *Wnt2* gene was upregulated in CTCs and metastases in a pancreatic cancer mouse model, using bulk RNA-seq and later single-cell RNA-ISH [13]. Upregulation of *Wnt2* in pancreatic cancer cells resulted in an increase of anchorage-independent sphere formation and metastatic propensity. The role of *Wnt2* in suppressing anoikis—programmed cell death in epithelial cells induced by detachment from the extracellular matrix—is through a non-canonical Wnt pathway mediated by *Tak1*, a serine/threonine protein kinase responsible for transcriptional regulation and apoptosis. In a follow-up study using scRNA-seq in pancreatic CTCs from both a mouse model and human patients, high expressions of stromal-derived extracellular matrix (ECM) proteins in CTCs, when compared to matching primary tumors, were discovered [8]. Knocking down the expression of SPARC, an EMC protein, suppressed cell migration and invasiveness, suggesting that the abnormal expression of stromal ECM proteins in CTCs could contribute to their metastatic spread to distant organs [8].

It has long been suspected that clusters of CTCs have higher metastatic potential than single CTCs. One group noted the association of CTC clusters with a worse prognosis in breast and prostate cancer patients and validated this hypothesis with cell line models [14]. To investigate the molecular drivers, they utilized scRNA-seq to analyze single CTCs and clustered CTCs from breast cancer patients and uncovered a list of differentially expressed genes including plakoglobin, which was implicated in cluster formation. Suppressing plakoglobin levels disrupted cluster formation and significantly suppressed the metastatic potential of those cells [14]. In a recent study, scRNA-seq analysis was applied to CTCs isolated from breast cancer patients with progressive metastatic lesions in bones or visceral organs [15]. The analysis discovered various enriched signaling pathways, including activated androgen receptor (AR) signaling in bone metastases. These patients have a correspondingly longer aromatase inhibitor (AI) treatment than patients with progressive visceral metastases [15]. This exciting finding suggests the role of AR signaling in promoting bone metastasis under the selective pressure of prolonged AI treatment, pointing

to a potential therapeutic opportunity to use AR targeted therapies that are already implemented in prostate cancers. These studies have demonstrated how single-cell transcriptomic analysis can facilitate the discovery of metastatic mechanisms.

scRNA-seq has also provided crucial insights on CTC transcriptional heterogeneity and its contribution to therapy resistance mechanisms. It has been discovered that transcriptional heterogeneity is present in CTCs from genetically engineered pancreatic cancer mouse models [8]. In addition, many epithelial CTCs express mesenchymal markers at various levels, which is consistent with a similar finding shown in breast cancer patient CTCs based on a multiplex fluorescent RNA-ISH assay [16]. The expression of these mesenchymal markers could contribute to the cancer stem cell-like characteristics [17] and resistance to various therapies [16, 18, 19]. In human patients, transcriptional heterogeneity is even more pronounced. scRNA-seq performed on prostate cancer CTCs demonstrated tremendous heterogeneity in transcriptomes, which could contribute to the various resistance mechanisms for AR-targeted therapies [9]. Transcriptional analyses between CTCs from patients with anti-androgen therapy (enzalutamide) resistance and naïve patients manifested two different resistance mechanisms: activation of glucocorticoid receptor (GR) and non-canonical Wnt signaling. Both pathways co-existed in CTCs to various degrees in different patients, including some CTCs from the treatment naïve group, pointing to the challenge of treating cancers with heterogeneous transcriptomes [9]. Similarly, another study showed heterogeneous levels of *SPINK1* and *BIRC5*—two genes associated with aggressive castration-resistant prostate cancers—in CTCs isolated from different prostate cancer patients [20].

These studies have shown that scRNA-seq can be used in CTCs as a powerful tool to discover novel tumor progression markers, which can ultimately guide personalized patient care. However, current scRNA-seq techniques have many limitations. Prior publications [8, 9, 20] have demonstrated that the success rate of overall amplification and library preparation is less than 60%, due to sample handling processes and the delicacy of CTCs [21]. In addition, single-cell experiments are highly prone to input loss. It is estimated that about 10–20% of transcripts are reverse transcribed during library preparation, thus favoring highly expressed genes over lesser expressed genes [22]. Other emerging techniques, such as RNA SPOTS [23] and sequential FISH [24], can possibly overcome this problem, due to their sensitivity for lesser expressed genes. Finally, CTCs that enter the bloodstream can potentially derive from various clones, or can be at different stages of replication, possibly confounding the analysis of expression patterns [8]. However, regardless of its limitations, scRNA-seq has become one of the most advanced single-cell methods and continues to improve. In parallel, CTC isolation is also evolving. In addition, studying numerous CTCs isolated from one patient or mouse model across multiple time points can overcome some of the technical limitations and assist in interpreting various findings.

Genetic Alterations

Profiling single cells from tumors can identify novel genetic alterations not found in bulk studies, revealing intra-tumor heterogeneity and the history of clonal evolution [25]. Most single-cell genetic studies have investigated copy number variations (CNVs) and gene

mutations. Mutational analyses are mainly accomplished with targeted exome sequencing and refer to single nucleotide point mutations or small insertions and deletions (INDELs) in exons. For example, one group deduced crucial information about genetic diversity and clonal history in primary and metastatic tumors [26]. Two breast cancer studies examined the evolutionary dynamics of CNVs in individual tumor cells using single-cell sequencing [26, 27]. Both studies demonstrated that complex aneuploid CNVs were attained early in tumor evolution, in punctuated bursts, followed by stable clonal expansions that eventually formed a tumor mass. Point mutations, on the other hand, accumulated gradually over a long period of time, expanding clonal diversity. In a recent study, topological imaging and single-cell genomic analysis were combined in the study of 10 synchronous breast cancer patients with both ductal carcinoma *in situ* (DCIS) and invasive ductal carcinoma (IDC) regions [28]. Based on the finding that most genetic alterations evolved in the *in-situ* regions prior to invasion, a multiclonal invasion model was proposed. In addition, another study has used single-cell CNV and mutational analyses in primary and metastatic colorectal cancer tissues from two patients and revealed monoclonal and polyclonal seeding of metastasis [29]. In this study, the results suggest a late dissemination model in these two patients because of many shared genetic alternations between primary and metastatic tumors.

Compared to the static time point analysis of primary and metastatic tumor samples, analyzing genetic alterations in single CTCs can reveal inter- and intra-patient heterogeneity and its association with therapeutic response in real-time. Similar to the above-mentioned findings in solid tissues, a study used Multiple Annealing and Looping Based Amplification Cycles (MALBAC) for exome sequencing and copy number profiling of single CTCs from seven patients with metastatic lung adenocarcinoma (ADC) and showed that CNVs in CTCs were highly stable [30]. Intriguingly, the CNV profiles from ADC and small-cell lung cancer (SCLC) subtypes were drastically different, whereas the global patterns of CNVs in different patients from the same cancer subtype showed a surprising conservation. In addition, CTC profiles were more like those from metastatic tumors versus primary tumors, including one patient who switched from an ADC profile in the primary tumor to an SCLC profile in a liver metastasis. Treatment with an SCLC standard regimen of etoposide and cisplatin resulted in a dramatic response in this patient, confirming the diagnostic value of CTCs in clinical management. On the other hand, it was also demonstrated that there was significant heterogeneity in point mutations and INDELs among single CTCs from a single patient, although a substantial number of point mutations (59%) from the primary and metastatic tumors were also detected in CTCs in this study.

Other studies of various cancers have reported CNV differences during disease progression. Studies of CTCs from SCLC patients have demonstrated clear patient-specific CNV profiles, while also detecting a few cases of intra-patient heterogeneity when comparing single CTCs to each other and to their complementary CTC-derived xenografts (a newly established “avatar” model for evaluating therapeutic responses of SCLC) [31]. Another study performed CNV sequencing via whole genome amplification in single CTCs isolated across four time points from a metastatic prostate cancer patient treated with chemotherapy and abiraterone [32]. The authors identified three different CNV subclones, each demonstrating various degrees of amplification in genes that potentially mediate therapeutic resistance, such as the *AR* and *MYC* genes. Additionally, they identified a specific CTC population that

showed genetic similarity to a retrospective biopsy taken from a bone metastasis, suggesting the origin for this CTC clone. In another study, a deletion of the *PTEN* gene in all CTCs from one patient was identified, and a unique clone with *BRAF* amplification was found in another patient [33]. Both a *PTEN* deletion and *BRAF* amplification have been described as resistant mechanisms to BRAF inhibitor therapy in melanoma patients [33]. Thus, by performing single-cell CNV analysis at different time points during therapy, this study has identified unique clones and underlying genetic alterations that arise during the selective pressures of therapy. Furthermore, CNV data from single CTCs was investigated to determine if it could distinguish chemosensitive from chemorefractory SCLC patients [34]. The scientists derived 16 CNV baseline profiles that demonstrated segregation of chemosensitive and chemorefractory status. This study was the first to identify potential genomic biomarkers in CTCs that indicate the chemosensitivity of an individual's SCLC before the start of treatment. Functional studies to identify the genetic mechanisms or regulatory regions in these locations would be even more informative for guiding effective chemotherapies. These studies have shown the power of using single-cell CNV in CTCs for the discovery of novel mechanisms of resistance, as well as potential biomarkers and therapeutic targets for future research.

The accurate assessment of point mutations in single CTCs is particularly challenging due to technical issues: amplification bias introduced by whole genome amplification, sequencing errors [35], and the sparseness and fragility of CTCs in a vial of blood. One group has tackled this issue by developing improved computational methods, sequencing multiple independent libraries of CTCs, and focusing on common mutations in primary and metastatic tumors [36]. Relying on these advancements, they have performed exome sequencing in single CTCs and compared their results with multiple regions of the primary tumor. Interestingly, there is obvious heterogeneity among various regions of the primary tumor and CTCs resemble one particular region, suggesting that they share a more recent common ancestor than others [36]. In addition, novel single-cell whole genome sequencing and computational methods, such as AneuFinder (a bioinformatic package that enables the detection of CNVs in single-cell sequencing data), will help address some of these challenges [37].

Aside from these two reports, most genomic studies are limited to targeted exome sequencing of a defined list of genes in a pooled population or a few exons of specific genes in single CTCs. Deep exome sequencing of a panel of 1000 cancer genes in a population of pure breast CTCs expanded *ev vivo* detected acquired mutations in CTCs shed from metastases, including the hotspot mutations in the ligand binding region of the estrogen receptor (ER), in addition to the driver mutations shared by the primary and metastatic tumors [38]. Even though this analysis is from pooled CTC lines, the allele frequency of various mutations suggests their polyclonal origin in certain patients. The significance of each clone in disease progression and treatment response can be investigated in depth due to the availability of the CTC lines. In another study the heterogeneity of *PIK3CA* mutations within single CTCs isolated from individual metastatic breast cancer patients was investigated [39]. It was found that different patients displayed various mutations within *PIK3CA*, and that one patient displayed three different *PIK3CA* variants in distinct single CTCs but demonstrated *PIK3CA* as wild-type in bulk samples. A different study revealed

that *KRAS* mutations were detected in 5 of 15 CTCs from one patient and *PIK3CA* mutations in 14 of 36 CTCs from four patients [40]. Furthermore, molecular characterization of single CTCs demonstrated considerable intra- and inter-patient heterogeneity of genetic alterations in *EGFR*, *KRAS*, and *PIK3CA*, possibly explaining the variable response rates to EGFR inhibition in patients with colorectal cancer. These studies have shown the value of single CTC analysis for specific genetic alterations that have clear clinical indications.

Even though whole genome sequencing is promising, amplifying the whole genome in order to develop a workable library for analysis can potentially result in a number of technical errors, including low physical sequencing coverage, non-uniform coverage, allelic dropout events, false-positive errors, and false-negative results due to insufficient coverage—the technical details for these errors were extensively discussed in other reviews [35, 41]. Special precaution must be taken in the post-processing of data in order to avoid the dilemma of calling every discovered variant heterogeneous [41]. In addition, *in-situ* sequencing has been developed even on samples fixed on slides [42]. This approach decreases the complexity of sample handling processes and the cell damage or loss that can occur during CTC enrichment and isolation [21].

DNA Methylation

DNA methylation plays an essential role in regulating many normal cell processes, and aberrant DNA methylation leads to the development of many diseases, especially cancer [43]. DNA methylation involves the addition of a methyl group to a cytosine (C), which is coupled with a guanine (G). Most CpG dinucleotides in the genome are methylated, but there are CpG islands—regions of DNA greater than 500 base pairs with a GC content equal to or greater than 55%—that are normally found unmethylated [44]. This unmethylated state is associated with the potential for gene expression. CpG islands represent about 70% or more of promoters within the genome and are correlated with genes that are essential for cell homeostasis, such as tumor suppressors [45]. Cancer tissues often show global hypomethylation and focal hypermethylation at promoters of tumor suppressor genes [43]. In addition, there are regions called partially methylated domains (PMDs)—megabase domains with lower-than-average-methylation—that are linked to gene repression and contain repressive histone marks. PMDs correspond to lamina-attachment domains—regions of DNA that bind to the nuclear wall—and are associated with gene poor regions, but little is known about their composition at the single-cell level [46]. Furthermore, methylation in certain enhancers impacts the expression of a cohort of genes [47]. Due to the extent of alterations occurring in cancer, DNA methylation is considered a typical hallmark of cancer. Currently, whole genome bisulfite sequencing (WGBS) is the gold standard to study methylation up to the single nucleotide level [48, 49], and it offers coverage for more than 90% of the approximately 28 million CpGs in the human genome [50].

There have been a few studies in which DNA methylation has been investigated in a population of CTCs. One study used a methylation array that interrogates 27,000 CpGs and found that there was hypermethylation in genes correlated with apoptosis, angiogenesis, and the VEGF pathway in CTCs when compared to non-matched primary tumors [51]. However, this study investigated less than 1% of all CpGs instead of globally. Another study involved

the use of pyrosequencing (sequencing by synthesis) and high-resolution melt analysis (amplicon melting analysis) to investigate the methylation of the *c-Met* and *Hgf* promoters in mouse derived CTC lines [52]. They found that *Hgf* and *c-Met* overexpression in the CTC lines were correlated with hypomethylation at their promoters. Both studies revealed epigenetic regulations occurring in CTCs, but the whole methylome (the DNA methylation patterns of all cytosines) has yet to be revealed. WGBS can be applied to a population of cells, but the underlying problem of investigating rare species persists.

For instance, most CTC isolating technologies result in the inclusion of contaminating leukocytes in the product, thus the methylome analysis from a population of cells in the product can lead to biased data. Single-cell WGBS analysis of CTCs eliminates the problem of contaminating cells and has the advantage of revealing the heterogeneity of methylomes and their association with disease progression.

There have been two recent single-cell methylation studies applied to single CTCs. One investigated the overall methylation levels of both DNA and RNA through mass spectrometry [53]. Another analyzed a few candidate epithelial-mesenchymal transition genes through single-cell multiplexed-agarose-embedded bisulfite sequencing [54]. Although both studies demonstrated that heterogeneity exists at the single-cell level, their methods only interrogated the overall methylation level or a few genes at a time, respectively. The excessive degradation caused by sodium bisulfite on DNA had previously restricted the possibility of applying WGBS to very low-input samples. Recently, there have been three published protocols that have achieved single-cell WGBS (scWGBS) in mouse ESCs and hepatocytes [50, 55, 56]. scWGBS enables the investigation of some unresolved questions in methylomes, such as whether PMDs exist at the single-cell level. It is also unknown whether PMDs are heterogeneous within a population of CTCs and whether the small number of genes within PMDs could contribute to therapeutic resistance and metastasis. In addition, the analysis of enhancer regions at the single-cell level could reveal potential heterogeneous clones due to the regulation of methylated enhancers.

Even though scWGBS is potentially possible in single CTCs, full genome-wide coverage is still challenging. As of now, scWGBS interrogates about 48% of all CpGs genome-wide [50]. Thus, allele-specific and strand-specific differences in methylation are difficult to detect, due to the low probability of covering both alleles and both strands for most CpG sites [57]. Genome-wide coverage may be achievable with a technique that aims to investigate all CpGs, CpG islands, enhancers, and PMDs. Despite the current technological limitations, future advances may lead to better CpG coverage at the single-cell level.

Chromatin Organization and Protein Interactions

In cancer, alterations in chromatin structure frequently involve gene silencing at nucleosome-dense regions across promoters [58]. Recent studies in lung and colon cancer have shown that the transformation of a tumor cell may originate with inappropriate regulation of nucleosome occupancy [58]. Consequently, this increases the binding of various transcription factors that alter gene regulation [58]. Genomic studies of nucleosome position, such as Formaldehyde-Assisted Isolation of Regulatory Elements (FAIRE-seq),

DNase I hypersensitive (DNase-seq) and Assay for Transposase-Accessible Chromatin sequencing (ATAC-seq), have been performed in cancer cells, but none of these techniques have been applied to CTCs due to the huge input of material required. Lately, single-cell studies investigating nucleosome positioning have become possible. Through the analysis of DNase-seq in individual tumor and normal cells from thyroid cancer patients, thousands of tumor-specific DNase-seq regions associated with promoters and enhancers that are critically involved in cancer development, have been discovered [59]. Additionally, single-cell ATAC-seq (scATAC-seq) in a myeloid leukemia cell line uncovered variance in accessibility associated with specific *trans*-factors and *cis*-elements [60].

Chromosomal conformation refers to the spatial organization of chromosomes that is responsible for gene regulation. Alterations in chromosomal conformation contribute to the biological behavior of different tumors, especially prostate cancer, breast cancer, and hematologic neoplasms [61]. Recently, a single-cell Hi-C (scHi-C) technique was used to construct 3D structures of a mouse ESC genome [62]. The study found that topological associated domains and loops varied extensively between cells. However, there is consistency across different cells in the genome-wide organization of lamina associated domains, A (active) and B (inactive) compartments, active enhancers and promoters, suggesting an important role of these regions in chromosome and genome folding.

DNA and protein interactions are commonly evaluated by chromatin immunoprecipitation sequencing (ChIP-seq), which typically requires $\sim 10^5$ cells [63]. A recent single-cell ChIP-seq (scChIP-seq) study combined microfluidics (encapsulating single cells based on microscale behavior of fluids) and DNA barcoding (short unique DNA sequences as barcodes for each sample) with sequencing to analyze mouse ESCs—with the limitation of very sparse unique reads per cell [64]. Nevertheless, by analyzing thousands of cells, the authors could distinguish variations in epigenetic states in distinct sub-populations of ESCs, which is not possible via bulk analysis.

Single-cell methods for chromatin accessibility, chromosome conformation, and DNA-protein interactions are still at an early stage with many limitations. scATAC-seq calls for less sample preparation than other methods and does not require the sonication or phenol extraction needed for FAIRE seq, or the sensitive enzyme digestion necessary for DNase-seq [65]. However, the most recent scATAC-seq method allows the mapping of only 70,000 reads per cell [60]. Single-cell DNase-seq can improve the mapping to 300,000 reads per cell, but with a 2% mapping efficiency [59, 66]. In addition, contaminating mitochondrial reads in ATAC-seq can potentially lead to low-coverage data sets [67]. Moreover, in scHi-C analysis, mitotic cells have shown disappearance of topologically associating domains and chromosomal A(active) and B (inactive) compartments, showing dramatic reorganization of the genomic structures during cell division [62]. Furthermore, the current scChIP-seq technique can only derive 1,000 unique reads per cell [64]. Despite the challenges of these techniques, the possibility of researching genomic organization and accessibility at the single-cell level within CTCs raises the promise of investigating variability of epigenetic regulation during the course of cancer treatment.

Concluding Remarks and Future Perspectives

CTCs are a unique resource for the real-time monitoring of active tumor biology in cancer patients. Because CTCs are rare, every cell counts. Single-cell sequencing analysis is particularly valuable in CTC research. RNA-seq and DNA-seq in CTCs have already facilitated important discoveries of intra-patient heterogeneity and potential mechanisms for treatment resistance, but much remains to be done (Table 1) (see Outstanding Questions). In addition, investigating the epigenetic regulation of single CTCs will be exciting since epigenetic mechanisms play a crucial role in drug resistance [68]. A recent study has shown that during *in vitro* expansion, individual CTCs can switch their HER2 expression to create an equilibrium of HER2 positive and negative subpopulations, suggesting an underlying epigenetic regulation [69]. These two distinct populations and the ability to convert between phenotypes create a challenge for targeting one population with a single therapy. Even though there are limitations associated with current techniques for analyzing epigenetic regulation in single cells, we expect advances in both library preparation and statistical analysis through the years. In addition, the potential of analyzing *ex vivo* cultured CTCs could help resolve some of the issues, although the *in vitro* expansion of CTCs poses its own challenges. Furthermore, analyzing a single CTC with many different types of modality would be extremely advantageous. Recent studies have been performed in single cells to analyze genomes and transcriptomes [70], DNA methylomes and transcriptomes [71], DNA methylome, chromatin accessibility, and nucleosome phasing [72]—or even genomes, transcriptomes and DNA methylomes all together [73]. Thus, in the coming years, the ability to evaluate the full spectrum of regulatory machinery in a single CTC may become a reality. There is no doubt that analyzing CTCs at the single cell level has shed light on tumor heterogeneity and therapeutic resistance mechanisms. Elucidating the phenotypic regulation during the cascade of metastasis or the course of therapeutic response in single CTCs remains a crucial question for the near future.

Outstanding Questions

- Which clones in CTCs are the most aggressive and what are their molecular properties that we need to monitor?
- How early can we detect metastasis-initiating CTCs?
- Is the metastasis-initiating trait of CTCs inherited in certain clones or induced during the adaptation to a new microenvironment?
- Can certain properties in CTCs be predictive for metastatic organ tropism?
- How do CTCs evade immune surveillance while swimming across a sea of immune cells?
- Are there unique properties of CTCs that can be targeted?

Acknowledgments

We apologize to colleagues whose studies were not incorporated in this review due to space limitations. We are grateful to Ms. C. Lytal for assisting us with the manuscript editing. Research in the M.Y. laboratories is funded by

the NIH K22 award (K22 CA175228-01A1), NIH DP2 award (DP2 CA206653), Donald E. and Delia B. Baxter Foundation, Stop Cancer Foundation, Wright Foundation, and the Pew Charitable Trusts and the Alexander & Margaret Stewart Trust (Pew-Stewart Scholar for Cancer Research) to M.Y., and the NIH F31 Fellowship award (F31 CA213970) to V.O.

References

1. Xue J, et al. Interrelationships of circulating tumor cells with metastasis and thrombosis: role of microRNAs. *Curr Pharm Des.* 2014; 20(33):5298–308. [PubMed: 24479797]
2. Kim MY, et al. Tumor self-seeding by circulating cancer cells. *Cell.* 2009; 139(7):1315–26. [PubMed: 20064377]
3. Yu M, et al. Circulating tumor cells: approaches to isolation and characterization. *J Cell Biol.* 2011; 192(3):373–82. [PubMed: 21300848]
4. Bedard PL, et al. Tumour heterogeneity in the clinic. *Nature.* 2013; 501(7467):355–64. [PubMed: 24048068]
5. Alix-Panabieres C, Pantel K. Circulating tumor cells: liquid biopsy of cancer. *Clin Chem.* 2013; 59(1):110–8. [PubMed: 23014601]
6. Alix-Panabieres C, Pantel K. Challenges in circulating tumour cell research. *Nat Rev Cancer.* 2014; 14(9):623–31. [PubMed: 25154812]
7. Dive C, Brady G. SnapShot: Circulating Tumor Cells. *Cell.* 2017; 169(1):176.
8. Ting DT, et al. Single-cell RNA sequencing identifies extracellular matrix gene expression by pancreatic circulating tumor cells. *Cell Rep.* 2014; 8(6):1905–1918. [PubMed: 25242334]
9. Miyamoto DT, et al. RNA-Seq of single prostate CTCs implicates noncanonical Wnt signaling in antiandrogen resistance. *Science.* 2015; 349(6254):1351–6. [PubMed: 26383955]
10. Miyamoto DT, et al. Androgen receptor signaling in circulating tumor cells as a marker of hormonally responsive prostate cancer. *Cancer Discov.* 2012; 2(11):995–1003. [PubMed: 23093251]
11. Lovly CM, Salama AK, Salgia R. Tumor Heterogeneity and Therapeutic Resistance. *Am Soc Clin Oncol Educ Book.* 2016; 35:e585–93. [PubMed: 27249771]
12. Ramskold D, et al. Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. *Nat Biotechnol.* 2012; 30(8):777–82. [PubMed: 22820318]
13. Yu M, et al. RNA sequencing of pancreatic circulating tumour cells implicates WNT signalling in metastasis. *Nature.* 2012; 487(7408):510–3. [PubMed: 22763454]
14. Aceto N, et al. Circulating tumor cell clusters are oligoclonal precursors of breast cancer metastasis. *Cell.* 2014; 158(5):1110–1122. [PubMed: 25171411]
15. Aceto N, et al. AR Expression in Breast Cancer CTCs Associates with Bone Metastases. *Mol Cancer Res.* 2018
16. Yu M, et al. Circulating breast tumor cells exhibit dynamic changes in epithelial and mesenchymal composition. *Science.* 2013; 339(6119):580–4. [PubMed: 23372014]
17. Mani SA, et al. The epithelial-mesenchymal transition generates cells with properties of stem cells. *Cell.* 2008; 133(4):704–15. [PubMed: 18485877]
18. Fischer KR, et al. Epithelial-to-mesenchymal transition is not required for lung metastasis but contributes to chemoresistance. *Nature.* 2015; 527(7579):472–6. [PubMed: 26560033]
19. Zheng X, et al. Epithelial-to-mesenchymal transition is dispensable for metastasis but induces chemoresistance in pancreatic cancer. *Nature.* 2015; 527(7579):525–530. [PubMed: 26560028]
20. Cann GM, et al. mRNA-Seq of single prostate cancer circulating tumor cells reveals recapitulation of gene expression and pathways found in prostate cancer. *PLoS One.* 2012; 7(11):e49144. [PubMed: 23145101]
21. Zhu Z, et al. Progress and challenges of sequencing and analyzing circulating tumor cells. *Cell Biol Toxicol.* 2017
22. Kolodziejczyk AA, et al. The technology and biology of single-cell RNA sequencing. *Mol Cell.* 2015; 58(4):610–20. [PubMed: 26000846]
23. Eng CL, et al. Profiling the transcriptome with RNA SPOTs. *Nat Methods.* 2017; 14(12):1153–1155. [PubMed: 29131163]

24. Lubeck E, Cai L. Single-cell systems biology by super-resolution imaging and combinatorial labeling. *Nat Methods*. 2012; 9(7):743–8. [PubMed: 22660740]
25. Mazor T, et al. Intratumoral Heterogeneity of the Epigenome. *Cancer Cell*. 2016; 29(4):440–451. [PubMed: 27070699]
26. Navin N, et al. Tumour evolution inferred by single-cell sequencing. *Nature*. 2011; 472(7341):90–4. [PubMed: 21399628]
27. Wang Y, et al. Clonal evolution in breast cancer revealed by single nucleus genome sequencing. *Nature*. 2014; 512(7513):155–60. [PubMed: 25079324]
28. Casasent AK, et al. Multiclonal Invasion in Breast Tumors Identified by Topographic Single Cell Sequencing. *Cell*. 2018; 172(1–2):205–217 e12. [PubMed: 29307488]
29. Leung ML, et al. Single-cell DNA sequencing reveals a late-dissemination model in metastatic colorectal cancer. *Genome Res*. 2017; 27(8):1287–1299. [PubMed: 28546418]
30. Ni X, et al. Reproducible copy number variation patterns among single circulating tumor cells of lung cancer patients. *Proc Natl Acad Sci U S A*. 2013; 110(52):21083–8. [PubMed: 24324171]
31. Hodgkinson CL, et al. Tumorigenicity and genetic profiling of circulating tumor cells in small-cell lung cancer. *Nat Med*. 2014; 20(8):897–903. [PubMed: 24880617]
32. Dago AE, et al. Rapid phenotypic and genomic change in response to therapeutic pressure in prostate cancer inferred by high content analysis of single circulating tumor cells. *PLoS One*. 2014; 9(8):e101777. [PubMed: 25084170]
33. Ruiz C, et al. Limited genomic heterogeneity of circulating melanoma cells in advanced stage patients. *Phys Biol*. 2015; 12(1):016008. [PubMed: 25574741]
34. Carter L, et al. Molecular analysis of circulating tumor cells identifies distinct copy-number profiles in patients with chemosensitive and chemorefractory small-cell lung cancer. *Nat Med*. 2017; 23(1):114–119. [PubMed: 27869802]
35. Sabina J, Leamon JH. Bias in Whole Genome Amplification: Causes and Considerations. *Methods Mol Biol*. 2015; 1347:15–41. [PubMed: 26374307]
36. Lohr JG, et al. Whole-exome sequencing of circulating tumor cells provides a window into metastatic prostate cancer. *Nat Biotechnol*. 2014; 32(5):479–84. [PubMed: 24752078]
37. van den Bos H, et al. Single-cell whole genome sequencing reveals no evidence for common aneuploidy in normal and Alzheimer's disease neurons. *Genome Biol*. 2016; 17(1):116. [PubMed: 27246599]
38. Yu M, et al. Cancer therapy. Ex vivo culture of circulating breast tumor cells for individualized testing of drug susceptibility. *Science*. 2014; 345(6193):216–20. [PubMed: 25013076]
39. Pestrin M, et al. Heterogeneity of PIK3CA mutational status at the single cell level in circulating tumor cells from metastatic breast cancer patients. *Mol Oncol*. 2015; 9(4):749–57. [PubMed: 25539732]
40. Gasch C, et al. Frequent detection of PIK3CA mutations in single circulating tumor cells of patients suffering from HER2-negative metastatic breast cancer. *Mol Oncol*. 2016; 10(8):1330–43. [PubMed: 27491860]
41. Navin NE. Tumor evolution in response to chemotherapy: phenotype versus genotype. *Cell Rep*. 2014; 6(3):417–9. [PubMed: 24529750]
42. Lee JH, et al. Fluorescent in situ sequencing (FISSEQ) of RNA for gene expression profiling in intact cells and tissues. *Nat Protoc*. 2015; 10(3):442–58. [PubMed: 25675209]
43. Esteller M. Aberrant DNA methylation as a cancer-inducing mechanism. *Annu Rev Pharmacol Toxicol*. 2005; 45:629–56. [PubMed: 15822191]
44. Takai D, Jones PA. Comprehensive analysis of CpG islands in human chromosomes 21 and 22. *Proc Natl Acad Sci U S A*. 2002; 99(6):3740–5. [PubMed: 11891299]
45. Deaton AM, Bird A. CpG islands and the regulation of transcription. *Genes Dev*. 2011; 25(10):1010–22. [PubMed: 21576262]
46. Berman BP, et al. Regions of focal DNA hypermethylation and long-range hypomethylation in colorectal cancer coincide with nuclear lamina-associated domains. *Nat Genet*. 2011; 44(1):40–6. [PubMed: 22120008]

47. Blattler A, et al. Global loss of DNA methylation uncovers intronic enhancers in genes showing expression changes. *Genome Biol.* 2014; 15(9):469. [PubMed: 25239471]
48. Cokus SJ, et al. Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. *Nature.* 2008; 452(7184):215–9. [PubMed: 18278030]
49. Lister R, et al. Highly integrated single-base resolution maps of the epigenome in Arabidopsis. *Cell.* 2008; 133(3):523–36. [PubMed: 18423832]
50. Smallwood SA, et al. Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity. *Nat Methods.* 2014; 11(8):817–820. [PubMed: 25042786]
51. Friedlander TW, et al. Detection and characterization of invasive circulating tumor cells derived from men with metastatic castration-resistant prostate cancer. *Int J Cancer.* 2014; 134(10):2284–93. [PubMed: 24166007]
52. Ogunwobi OO, et al. Epigenetic upregulation of HGF and c-Met drives metastasis in hepatocellular carcinoma. *PLoS One.* 2013; 8(5):e63765. [PubMed: 23723997]
53. Huang W, et al. Determination of DNA and RNA Methylation in Circulating Tumor Cells by Mass Spectrometry. *Anal Chem.* 2016; 88(2):1378–84. [PubMed: 26707930]
54. Pixberg CF, et al. Analysis of DNA methylation in single circulating tumor cells. *Oncogene.* 2017; 36(23):3223–3231. [PubMed: 28068321]
55. Farlik M, et al. Single-cell DNA methylome sequencing and bioinformatic inference of epigenomic cell-state dynamics. *Cell Rep.* 2015; 10(8):1386–97. [PubMed: 25732828]
56. Gravina S, et al. Single-cell genome-wide bisulfite sequencing uncovers extensive heterogeneity in the mouse liver methylome. *Genome Biol.* 2016; 17(1):150. [PubMed: 27380908]
57. Clark SJ, et al. Genome-wide base-resolution mapping of DNA methylation in single cells using single-cell bisulfite sequencing (scBS-seq). *Nat Protoc.* 2017; 12(3):534–547. [PubMed: 28182018]
58. Druliner BR, et al. Comprehensive nucleosome mapping of the human genome in cancer progression. *Oncotarget.* 2016; 7(12):13429–45. [PubMed: 26735342]
59. Jin W, et al. Genome-wide detection of DNase I hypersensitive sites in single cells and FFPE tissue samples. *Nature.* 2015; 528(7580):142–6. [PubMed: 26605532]
60. Buenrostro JD, et al. Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature.* 2015; 523(7561):486–90. [PubMed: 26083756]
61. Jia R, et al. Novel insights into chromosomal conformations in cancer. *Mol Cancer.* 2017; 16(1):173. [PubMed: 29149895]
62. Stevens TJ, et al. 3D structures of individual mammalian genomes studied by single-cell Hi-C. *Nature.* 2017; 544(7648):59–64. [PubMed: 28289288]
63. Nakato R, Shirahige K. Recent advances in ChIP-seq analysis: from quality management to whole-genome annotation. *Brief Bioinform.* 2017; 18(2):279–290. [PubMed: 26979602]
64. Rotem A, et al. Single-cell ChIP-seq reveals cell subpopulations defined by chromatin state. *Nat Biotechnol.* 2015; 33(11):1165–72. [PubMed: 26458175]
65. Meyer CA, Liu XS. Identifying and mitigating bias in next-generation sequencing methods for chromatin biology. *Nat Rev Genet.* 2014; 15(11):709–21. [PubMed: 25223782]
66. Clark SJ, et al. Single-cell epigenomics: powerful new methods for understanding gene regulation and cell identity. *Genome Biol.* 2016; 17:72. [PubMed: 27091476]
67. Montefiori L, et al. Reducing mitochondrial reads in ATAC-seq using CRISPR/Cas9. *Sci Rep.* 2017; 7(1):2451. [PubMed: 28550296]
68. Sharma SV, et al. A chromatin-mediated reversible drug-tolerant state in cancer cell subpopulations. *Cell.* 2010; 141(1):69–80. [PubMed: 20371346]
69. Jordan NV, et al. HER2 expression identifies dynamic functional states within circulating breast cancer cells. *Nature.* 2016; 537(7618):102–106. [PubMed: 27556950]
70. Macaulay IC, et al. G&T-seq: parallel sequencing of single-cell genomes and transcriptomes. *Nat Methods.* 2015; 12(6):519–22. [PubMed: 25915121]
71. Angermueller C, et al. Parallel single-cell sequencing links transcriptional and epigenetic heterogeneity. *Nat Methods.* 2016; 13(3):229–232. [PubMed: 26752769]

72. Pott S. Simultaneous measurement of chromatin accessibility, DNA methylation, and nucleosome phasing in single cells. *Elife*. 2017;6.
73. Hou Y, et al. Single-cell triple omics sequencing reveals genetic, epigenetic, and transcriptomic heterogeneity in hepatocellular carcinomas. *Cell Res*. 2016; 26(3):304–19. [PubMed: 26902283]
74. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet*. 2009; 10(1):57–63. [PubMed: 19015660]
75. Boyle AP, et al. High-resolution mapping and characterization of open chromatin across the genome. *Cell*. 2008; 132(2):311–22. [PubMed: 18243105]
76. Buenrostro JD, et al. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods*. 2013; 10(12): 1213–8. [PubMed: 24097267]
77. van Berkum NL, et al. Hi-C: a method to study the three-dimensional architecture of genomes. *J Vis Exp*. 2010; (39)
78. Johnson DS, et al. Genome-wide mapping of in vivo protein-DNA interactions. *Science*. 2007; 316(5830):1497–502. [PubMed: 17540862]

Highlights

- Circulating tumor cells (CTCs) are a rare population of tumor cells in the blood stream and represent a mixture of heterogeneous cells shed from various regions of primary or metastatic tumors.
- Analyzing CTCs throughout the disease course of a cancer patient provides an alternative means for tumor biopsy in real-time.
- Traditional genome-wide studies depict an average of a complex population, whereas single-cell studies have the advantage of investigating heterogeneity and rare cell populations.
- Genome-wide single-cell RNA-Seq and DNA-seq performed in CTCs have provided crucial new insights into CTC heterogeneity and mechanisms of therapeutic resistance.
- Single cell analyses for DNA methylation, protein-DNA interactions, and the chromosomal organization have been applied to other cell types and remain to be done in CTCs.

Figure 1, Key Figure. Illustration of current genome-wide technologies and potential applications for single CTCs

Blood samples from cancer patients are processed and CTCs are isolated. DNA or RNA, or both, are isolated and processed by one or more of the following single-cell techniques: RNA-Seq, DNA-Seq, WGBS (whole genome bisulfite sequencing), DNase-Seq (DNase I hypersensitive Sequencing), ATAC-Seq (Assay for Transposase-Accessible Chromatin Sequencing), Hi-C Seq, ChIP-Seq (Chromatin Immunoprecipitation Sequencing). RNA-Seq depicts exomes and splicing probabilities. DNase-Seq and ATAC-Seq depicts nucleosomes on DNA. Hi-C Seq depicts an example of how DNA can be organized structurally within the nucleus. ChIP-Seq depicts a protein binding to DNA, which is captured by a targeting antibody. These techniques can manifest heterogeneous population of cells, personalize prognosis and treatment, enable therapy monitoring, manifest drug resistance mechanisms and lead to the discovery of novel regulatory mechanisms that may advance our understanding of cancer and make an impact on cancer treatments.

Table 1

Applied and potential genome-wide studies in bulk and single CTCs.

Genome-Wide Single-Cell Techniques	Approach	Performed in Bulk/Single CTCs?
RNA-seq	Transcriptome	Yes/Yes [6–11, 16]
DNA-seq	Genetic	Yes/Yes [25–30]
WGBS *	DNA Methylation	No/No
DNase-seq *	Nucleosome Position	No/No
ATAC-seq *	Nucleosome Position	No/No
HiC-seq *	3D Genome Organization	No/No
ChIP-seq *	DNA-Protein Interactions	No/No

* WGBS (Whole Genome Bisulfite Sequencing); DNase-seq (DNase I hypersensitive Sequencing); ATAC-seq (Assay for Transposase-Accessible Chromatin Sequencing); HiC-seq (a method that probes the 3D architecture of whole genomes by coupling proximity-based ligation with sequencing); ChIP-seq (Chromatin Immunoprecipitation Sequencing)